The Journal of Molecular Diagnostics, Vol. 14, No. 4, July 2012

http://dx.doi.org/10.1016/j.jmoldx.2012.02.007

Technical Advance

A Novel Long-Range PCR Sequencing Method for Genetic Analysis of the Entire *PKD1* Gene

Ying-Cai Tan,* Alber Michaeel,* Jon Blumenfeld,†‡ Stephanie Donahue,‡ Tom Parker,‡ Daniel Levine,‡ and Hanna Rennert*

From the Departments of Pathology and Laboratory Medicine and Medicine,† Weill Cornell Medical College, New York; and The Rogosin Institute,‡ New York, New York*

Genetic testing of *PKD1* and *PKD2* is useful for the diagnosis and prognosis of autosomal dominant polycystic kidney disease; however, analysis is complicated by the large transcript size, the complexity of the gene region, and the high level of gene variations. We developed a novel mutation screening assay for *PKD1* by directly sequencing long-range (LR) PCR products. By using this method, the entire *PKD1* coding region was amplified by nine reactions, generating product sizes from 2 to 6 kb, circumventing the need for specific PCR amplification of individual exons. This method was compared with direct sequencing used by a reference laboratory and the SURVEYOR-WAVE Nucleic Acid High Sensitivity Fragment Analysis System (Transgenomic) screening method for five patients with autosomal dominant polycystic kidney disease. A total of 53 heterozygous genetic changes were identified by LR PCR sequencing, including 41 (of 42) variations detected by SURVEYOR nuclease and all 32 variations reported by the reference laboratory, detecting an additional 12 intronic changes not identified by the other two methods. Compared with the reference laboratory, LR PCR sequencing had a sensitivity of 100%, a specificity of 98.5%, and an accuracy of 98.8%; compared with the SURVEYOR-WAVE method, it had a sensitivity of 97.1%, a specificity of 100%, and an accuracy of 99.4%. In conclusion, LR PCR sequencing was superior to the direct sequencing and screening methods for detecting genetic variations, achieving high sensitivity and improved intronic coverage with a faster turnaround time and lower costs, and providing a reliable tool for complex genetic analyses. *(J Mol Diagn 2012, 14:305-313; http://dx.doi.org/10.1016/j.jmoldx.2012.02.007)*

Autosomal dominant polycystic kidney disease (ADPKD) is the most common inherited kidney disease, affecting approximately 1 in 500 individuals in the United States and 12.5 million patients worldwide. It is characterized by bilateral kidney cyst development[1–4] and progressive chronic kidney disease, leading to end-stage renal disease.[5,6] ADPKD is caused by mutations in two genes, with the *PKD1* gene accounting for 75% to 85% of the cases and the *PKD2* gene responsible for the remainder of the cases.[7] Genetic testing plays an increasingly significant role in the diagnosis of patients with an unclear renal phenotype, particularly in the absence of a known family history, and in the evaluation of family members who are considering kidney donation to affected individuals.[8] However, genetic testing is complicated by the presence of *PKD1* homologous genes (HGs), the large size of *PKD1*, and marked allelic heterogeneity of the disease-associated mutations.[9,10]

The main strategies used for ADPKD genetic analysis are all PCR based, followed by sequencing of the corresponding exons. Although the entire *PKD2* gene and the single-copy region of the *PKD1* gene can be directly amplified by PCR from genomic DNA, the duplicated *PKD1* region requires nested PCR to amplify the true gene among the highly identical HGs.[11] The nested-PCR approach comprises two primary steps: a long-range (LR) PCR with primers located in the rare mismatch sites

Supported by a grant from the National Center for Research Resources (UL1RR024143), a component of the NIH and NIH Roadmap for Medical Research.

Accepted for publication February 22, 2012.

The contents of this article are solely the responsibility of the authors and do not necessarily represent the official view of the National Center for Research Resources or NIH.

CME Disclosure: The authors of this article and the planning committee members and staff have no relevant financial relationships with commercial interest to disclose.

Address reprint requests to Hanna Rennert, Ph.D., Department of Pathology and Laboratory Medicine, Weill Cornell Medical College, 525 E. 68th St., Room F701, New York, NY 10065. E-mail: har2006@med.cornell.edu.

Table 1. *PKD1* Primers for LR PCR

Fragment	Size (bp)	Exon	Forward primer	Reverse primer	Temperature (°C)
Gene1	2278	1	5′-CGCAGCCTTACCATCCACCT-3′	5′-TCATCGCCCCTTCCTAAGCA-3′	64
Gene2-7	4041	2-7	5′-CCCCGAGTAGCTGGAACTACAGTTAC-ACACT-3′	5′-CGTCCTGCTGTGCCAGAGGCG-3′	70
Gene8-12	3893	8-12	5′-ACGTCTGCGAGCTGCAGCCC-3′	5′-CTGCAGGGACAGGCGTCAGTGA-3′	70
Gene13-15	4391	13-15	5′-TGGAGGGAGGGACGCCAATC-3′	5′-GTCAACGTGGGCCTCCAAGT-3′	65
Gene15-21	5253	15-21	5′-ATCCCTGGGGGTCCTACCATCTCTTA-3′	5′-ACACAGGACAGAACGGCTGAGGCTA-3′	68
Gene22-26	3276	22-26	5′-ATGCTTAGTGAGGAGGCTGTGGG-GGTCCA-3′	5′-GCTTAAAGGGGAATGGCTTAAACCCG-3′	70
Gene27-34	3916	27-34	5′-CGGGTCACCGGTTGTGGCA-3′	5′-ATGAGGCTCTTTCCACAGACAAC-AGAGGTT-3′	70
Gene35-41	2632	35-41	5′-CAAGAGGCTCAAGAAACTGCCCG-3′	5′-GGGCTGTGGAAGCCGCCTA-3′	68
Gene42-46	2370	42-46	5′-GAGTAGTTCTCCAGGAGTGCCG-3′	5′-ATTCTGCCTGGCCCTCGGCCTT-3′	63

Table 2. LR PCR Amplification Conditions

PKD1 sequence	PCR conditions
Gene1	99°C for 5 minutes; followed by cooling down to 95°C for 2 minutes; and 35 cycles of 95°C for 1 minute, 62°C for 1 minute, and 72°C for 3 minutes, with a final extension step at 72°C for 10 minutes
Gene2-7 and Gene8-12	Touchdown protocol composed of an initial step of 95°C for 3 minutes; followed by 14 cycles of 95°C for 30 seconds, 74°C for 30 seconds, with a decrease of 0.5°C per cycle, and 68°C for 5 minutes; followed by 30 cycles of 95°C for 30 seconds, 67°C for 30 seconds, and 68°C for 5 minutes, with a final extension step of 72°C for 10 minutes
Gene13-15	95°C for 3.15 minutes, followed by 35 cycles of 95°C for 20 seconds and 68°C for 5 minutes, with a final extension step at 72°C for 7 minutes
Gene15-21	95°C for 3 minutes, followed by 35 cycles of 95°C for 30 seconds and 68°C for 4.5 minutes, with a final extension step at 72°C for 10 minutes
Gene22-26, Gene27-34, Gene35-41, and Gene42-46	94°C for 3 minutes; followed by 14 cycles of 94°C for 30 seconds, 66°C for 30 seconds, with a decrease of 0.5°C per cycle, and 68°C for 3.5 minutes; followed by 30 cycles of 95°C for 30 seconds, 59°C for 30 seconds, and 68°C for 3.5 minutes, with a final extension step of 72°C for 10 minutes

Table 3. *PKD1* Sequencing Primers

Primer name	Sequence	Genomic position (of 5′ first bp)	Distance to exon-intron junction	Exons covered
PKD1–exon1-F	5′-GCGTCGCTCAGCAGCAGGT-3′	2,185,830	−140 Ex1	1
PKD1–exon1-R	5′-GCCCGCGTCCTGCTTCCC-3′	2,185,383	93 Ex1	
PKD1–exon2-3-F	5′-GGGATGCTGGCAATGTGTGGGAT-3′	2,169,468	−89 Ex2	2-3
PKD1–exon2-3-R	5′-GGACCAACTGGGAGGGCAGAA-3′	2,169,038	77 Ex3	
PKD1–exon4-F	5′-GGCGGTGCTGTCAGGGTG-3′	2,168,948	−102 Ex4	4
PKD1–exon4-R	5′-CCAGAGAGGCCTTCCTGAGC-3′	2,168,582	95 Ex4	
PKD1–exon5A-F	5′-GAACAGCATGGGAGCCTGTGAGT-3′	2,168,561	−98 Ex5	5
PKD1–exon5A-R	5′-AGCCGGCCCAGCGGCATC-3′	2,168,033	Within an exon	
PKD1–exon5B-F	5′-GCCTGTCCCTCTGCTCCG-3′	2,168,262	Within an exon	5
PKD1–exon5B-R	5′-GTGTCAACGGTCAGTGTGGGC-3′	2,167,696	96 Ex5	
PKD1–exon6-F	5′-GTGTCTGCTGCCCACTCCC-3′	2,167,787	−124 Ex6	6
PKD1–exon6-R	5′-CTCCTTCCTCCTGAGACTCCC-3′	2,167,397	93 Ex6	
PKD1–exon7-F	5′-GCTGCTGTGAGGGTGGGAGGA-3′	2,167,120	−66 Ex7	7
PKD1–exon7-R	5′-TCCACCGCGGGCGCTCGGCA-3′	2,166,763	71 Ex7	
PKD1–exon8-F	5′-CTGGGCTGAGGAGGAGGG-3′	2,166,760	−115 Ex8	8
PKD1–exon8-R	5′-GGGCACAAGCAACATTAAGGCCC-3′	2,166,413	117 Ex8	
PKD1–exon9-F	5′-CCTCTTCCTGGGAAGTTCGGGT-3′	2,166,206	−87 Ex9	9
PKD1–exon9-R	5′-ACTCTGGTGGCCACAGGACCA-3′	2,165,895	98 Ex9	
PKD1–exon10-F	5′-GCAGGCAGTTGGGCATCTCTG-3′	2,165,697	−71 Ex10	10
PKD1–exon10-R	5′-GACCCTGGGCAGCAGACAG-3′	2,165,309	70 Ex10	
PKD1–exon11A-F	5′-GTGTGGCTGACGAAGCGGG-3′	2,165,035	−109 Ex11	11
PKD1–exon11A-R	5′-CCGTGGCGTTGGCACCAG-3′	2,164,455	Within an exon	
PKD1–exon11B-F	5′-CGCTATGAGGTCCGGGCAG-3′	2,164,641	Within an exon	11
PKD1–exon11B-R	5′-CCCTCACTGGGAAGCCAGG-3′	2,164,077	93 Ex11	
PKD1–exon12-F	5′-GGACTCTCCAGCCCGACG-3′	2,163,387	−94 Ex12	12
PKD1–exon12-R	5′-GCAGAGGTGAAGGTGGAGC-3′	2,163,088	74 Ex12	
PKD1–exon13-14-F	5′-GTGGAGGGAGGGACGCCAA-3′	2,163,037	−73 Ex13	13-14
PKD1–exon13-14-R	5′-GTCACAGTGAGGGCTGTTGGG-3′	2,162,230	111 Ex14	
PKD1–exon15A-F	5′-TTCTGCCGAGCGGGTGGG-3′	2,161,938	−64 Ex15	15
PKD1–exon15A-R	5′-CATGTCGAAGGTCCACGTGATGT-3′	2,161,427	Within an exon	

table continues

Table 3. *Continued*

Primer name	Sequence	Genomic position (of 5′ first bp)	Distance to exon-intron junction	Exons covered
PKD1–exon15B-F	5′-GACATGAGCCTGGCCGTGG-3′	2,161,516	Within an exon	15
PKD1–exon15B-R	5′-CCACCTCTGGCTCCACGCA-3′	2,161,027	Within an exon	
PKD1–exon15C-F	5′-CACGCGGAGCGGCACGTT-3′	2,161,121	Within an exon	15
PKD1–exon15C-R	5′-GGTGACCTCCGGACCCTC-3′	2,160,626	Within an exon	
PKD1–exon15D-F	5′-TCTGCTGTGGGCCGTGGG-3′	2,160,706	Within an exon	15
PKD1–exon15D-R	5′-CTGTACCGTGTGGTTGGTGGG-3′	2,160,215	Within an exon	
PKD1–exon15E-F	5′-ACAGCATCTTCGTCTATGTCCTG-3′	2,160,303	Within an exon	15
PKD1–exon15E-R	5′-GGTTCCCTGCCGTCATGGTG-3′	2,159,812	Within an exon	
PKD1–exon15F-F	5′-GGGCTGAGCTGGGAGACCT-3′	2,159,899	Within an exon	15
PKD1–exon15F-R	5′-GACAGCTGAGCCGGCAGC-3′	2,159,417	Within an exon	
PKD1–exon15G-F	5′-CTGTGGGCCAGCAGCAAGGT-3′	2,159,494	Within an exon	15
PKD1–exon15G-R	5′-CGTGCGGTTCTCACTGCCCA-3′	2,159,012	Within an exon	
PKD1–exon15H-F	5′-GACGTCACCTACACGCCCG-3′	2,159,095	Within an exon	15
PKD1–exon15H-R	5′-CCTCCCAGCGGTACTCAGTCT-3′	2,158,603	Within an exon	
PKD1–exon15I-F	5′-GATGCGGCGATCACAGCGCA-3′	2,158,688	Within an exon	15
PKD1–exon15I-R	5′-GGCCAGCCCTGGTGGCAA-3′	2,158,164	89 Ex15	
PKD1–exon16-F	5′-GGCCCGTCCTCAGTGCCT-3′	2,158,167	−134 Ex16	16
PKD1–exon16-R	5′-GCGGCCTCCACCAGCACTA-3′	2,157,790	94 Ex16	
PKD1–exon17-18-F	5′-GAAACCTGGAGTTTGGGAGCAGC-3′	2,157,047	−98 Ex17	17-18
PKD1–exon17-18-R	5′-TGACGTCACAGAGTCGGG-3′	2,156,344	55 Ex18	
PKD1–exon19-20-F	5′-GCACGGGTGAGTGCAGGC-3′	2,156,404	−99 Ex19	19-20
PKD1–exon19-20-R	5′-CCGGGATGAGCCCTCTGCAA-3′	2,155,768	98 Ex20	
PKD1–exon21-F	5′-AGTCGTGGGCATCTGCTGGC-3′	2,155,550	−75 Ex21	21
PKD1–exon21-R	5′-CAAGCTGCCCGTCTGCCCT-3′	2,155,240	83 Ex21	
PKD1–exon22-F	5′-CAGGTGAGGACCCGTGTAGAGA-3′	2,154,725	−82 Ex22	22
PKD1–exon22-R	5′-GGGAGGAGGGAGGCAGAG-3′	2,154,431	68 Ex22	
PKD1–exon23A-F	5′-GCACCTCGCTCTCTGCCC-3′	2,154,024	−128 Ex23	23
PKD1–exon23A-R	5′-GCCACCTTGGTGGAGACGG-3′	2,153,512	Within an exon	
PKD1–exon23B-F	5′-GGCTGCCACTTCTCCATCCC-3′	2,153,651	Within an exon	23
PKD1–exon23B-R	5′-GACACCCATGGAAGCCCTACG-3′	2,153,183	84 Ex23	
PKD1–exon24-F	5′-CGTGGCAGAGGGTGGGCT-3′	2,153,064	−93 Ex24	24
PKD1–exon24-R	5′-CTCGCTGCCTGCCGTCCC-3′	2,152,721	94 Ex24	
PKD1–exon25-26-F	5′-GGCTCTGAGACTGCGACATCCAA-3′	2,152,702	−68 Ex25	25-26
PKD1–exon25-26-R	5′-CTTGTTCTGACGCCTGCGACG-3′	2,151,964	98 Ex26	
PKD1–exon27-28-F	5′-GCTGAGATGACTTGCCTGGGATG-3′	2,150,644	−77 Ex27	27-28
PKD1–exon27-28-R	5′-ACTGCAGGAGGCCACGGG-3′	2,150,094	73 Ex28	
PKD1–exon29-30-F	5′-CTCCGTGGGAGGTTGGGCA-3′	2,150,142	−70 Ex29	29-30
PKD1–exon29-30-R	5′-CGCCTTTCCCTCTGGCTGC-3′	2,149,542	103 Ex30	
PKD1–exon31-32F	5′-CGGGCTCTGTCCTGTCTGC-3′	2,148,079	−94 Ex31	31-32
PKD1–exon31-32R	5′-CCCAGCAAGGACACGCAGC-3′	2,147,642	87 Ex32	
PKD1–exon33-34-F	5′-GGAAGCCCAGGGTGTCCGT-3′	2,147,595	−91 Ex33	33-34
PKD1–exon33-34-R	5′-CAGCCCTGCCCTGGCACC-3′	2,147,034	115 Ex34	
PKD1–exon35-F	5′-CAAGAGGCTCAAGAAACTGCCCG-3′	2,144,309	−98 Ex35	35-36
PKD1–exon36-R	5′-GAGAAGTACAGGGCTTCCAGCAA-3′	2,143,714	98 Ex36	
PKD1–exon37-F	5′-CTCGGCTGGGAGCCACTG-3′	2,143,832	−93 Ex37	37
PKD1–exon37-R	5′-GCCTTCTGAGGTGAGGAAAGGG-3′	2,143,439	106 Ex37	
PKD1–exon38-F	5′-CCACACCTGCCGCAGCCAT-3′	2,143,190	−96 Ex38	38
PKD1–exon38-R	5′-CAAAGGTATCTACACATGTCCAC-3′	2,142,856	99 Ex38	
PKD1–exon39-F	5′-GCCAGCAGGGCAGTGGGA-3′	2,142,698	−105 Ex39	39
PKD1–exon39-R	5′-CAGCTAGGGAGCAGGGCTGA-3′	2,142,385	96 Ex39	
PKD1–exon40-F	5′-GTGGCGCCGAACCAGAGC-3′	2,142,289	−100 Ex40	40-41
PKD1–exon41-R	5′-GGGCTGTGGAAGCCGCCTA-3′	2,141,678	104 Ex41	
PKD1–exon42-F	5′-CCTCAGCCACGCCTGCACT-3′	2,141,678	−80 Ex42	42
PKD1–exon42-R	5′-GGGTGAGACGCTGCCGGG-3′	2,141,338	86 Ex42	
PKD1–exon43-F	5′-CAGCGTCCCTCCCGCCCT-3′	2,141,213	−38 Ex43	43-44
PKD1–exon44-R	5′-CAGGAAGACACGAGCTGCGG-3′	2,140,578	97 Ex44	
PKD1–exon45-F	5′-GCTGGCCATCCTGGTAGGTGA-3′	2,140,687	−96 Ex45	45
PKD1–exon45-R	5′-GGACTTTGTGGCGGAACTGGG-3′	2,140,180	106 Ex45	
PKD1–exon46-F	5′-GGAGAGGGACACGCCCTG-3′	2,140,264	−69 Ex46	46
PKD1–exon46-R	5′-ATTCTGCCTGGCCCTCGGCCTT-3′	2,139,635	Within 3′UTR	

The genomic primer positions were according to the February 2009 human reference sequence (GRCh37/hg19).
Ex, exon.

distinguishing *PKD1* and the HGs and a nested PCR to amplify each individual exon. This step prevents amplification of the *PKD1* HGs that potentially could confound the analysis.[11] However, this process is highly labor intensive, susceptible to PCR contamination, and expensive. The number of sequencing reactions can be substantially reduced by mutation screening procedures[12]; however, these procedures do not eliminate the need for

a nested PCR. To improve genetic analysis of ADPKD and to avoid the use of a nested PCR, we have developed a new method for *PKD1* genetic testing by directly sequencing LR PCR fragments covering the entire *PKD1* gene. In this study, the accuracy of the new assay was evaluated using a panel of DNA samples that were previously analyzed by two different methods that have been previously validated.[12]

Materials and Methods

Study Subjects

Study subjects were participants in The Rogosin Institute ADPKD Data Repository (NIH clinicaltrials.gov website; *http://clinicaltrials.gov* identifier: NCT00792155, last accessed May 15, 2012). This is a single-center longitudinal study of genotype and phenotype characteristics of individuals with ADPKD. Study samples were obtained from five subjects from the repository who were randomly selected for analysis. All subjects underwent PKD genotyping by the Athena Diagnostics, Inc., reference laboratory (Worcester, MA) and by the Weill Cornell Medical Center Molecular Pathology Research Laboratory, New York, NY, using direct sequencing and SURVEYOR nuclease–WAVE screening, respectively. The study was approved by the Institutional Review Board Committees at Weill Cornell Medical Center (New York, NY) and The Rockefeller University (New York, NY). All subjects provided written informed consent.

PCR Amplification and Sequencing

Genomic DNA was extracted from peripheral blood lymphocytes using standard DNA extraction methods. The entire coding region, 5′ and 3′ untranslated regions, and exon-intron boundaries of *PKD1* were amplified in nine distinct LR PCRs (Gene1, Gene2-7, Gene8-12, Gene13-15, Gene15-21, Gene22-26, Gene27-34, Gene35-41, and Gene42-46), using PCR primers (Sigma-Genosys Ltd, St Louis, MO) anchored in either the rare mismatched region with the HGs or the single-copy region of *PKD1*. The LR PCR primer sequences are shown in Table 1. LR PCR was performed using the GeneAmp High Fidelity PCR System (Applied Biosystems, Foster City, CA), as previously described.[12] Briefly, 60 ng of genomic DNA was amplified in a final volume of 25 μL, containing 200 μmol/L deoxyribonucleotide triphosphate, 0.2 μmol/L of each primer, 0.5 mol/L betaine, 5% dimethyl sulfoxide (except for exon 1 with an extremely high content of GC, for which 10% dimethyl sulfoxide was used), manufacture's supplied buffer, and 2 U of enzyme. The LR PCR products were amplified separately using the Biometra T-3000 thermalcycler (Biometra GmbH, Goettingen, Germany), as described in Table 2. The LR PCR products of *PKD1* were then purified with the Qiaquick PCR purification kit (Qiagen Inc., Germantown, MA), quantified, and sequenced with 45 pairs of walking primers located at least 50 bp away from intron-exon junctions, using Big Dye Terminator Chemistry with AmpliTaq-FS DNA Polymerase (Applied Biosystems, Carlsbad, CA) on an ABI 3100 Genetic Analyzer (Table 3). Sequencing data (ABI file) were analyzed by Mutation Surveyor software version 4.0 (SoftGenetics, State College, PA) for automatic variation calling first, followed by careful inspection of the electropherograms for quality assurance purposes.

Assay Analytical Characteristics

An evaluation of the assay analytical characteristics was performed using only variants located in regions sequenced by all three methods used and compared herein, specifically exons and nearby intron sequences. The new assay was assessed for sensitivity, specificity, and accuracy using the following formulas: sensitivity = number of true positive/(number of true positive + number of false negative), specificity = number of true negative/(number of true negative + number of false positive), and accuracy = (number of true positive + number of true negative)/(number of true positive + number of false positive + number of false negative + number of true negative). The 95%

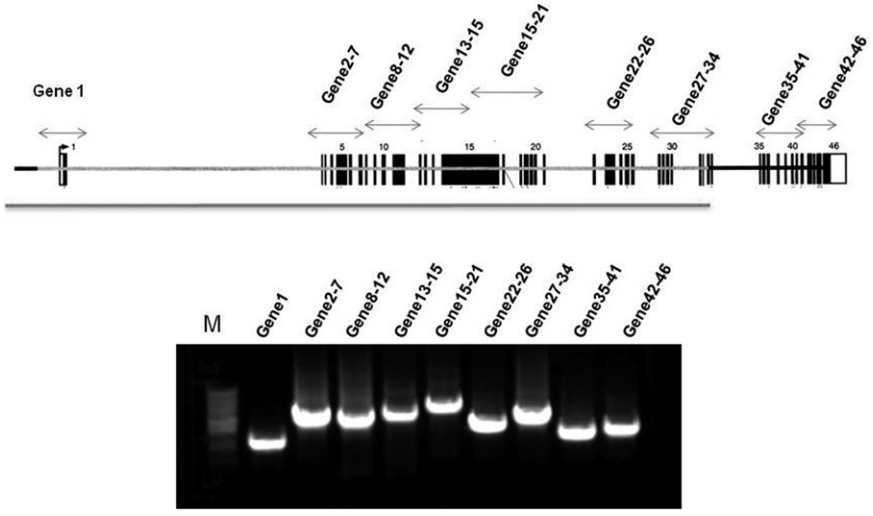

Figure 1. Top: Map of the *PKD1* gene showing the position of the nine pairs of primers used for LR PCR amplification of the entire gene. The structure of the *PKD1* gene is shown, with genes 1 to 46 indicated. Black line indicates the duplicated region of *PKD1*. **Bottom:** An example of LR PCR amplification of the DNA fragments (2 to 6 kb long) containing all exons and flanking regions of *PKD1* analyzed on a 0.5% agarose gel with ethidium-bromide staining. M, size marker.

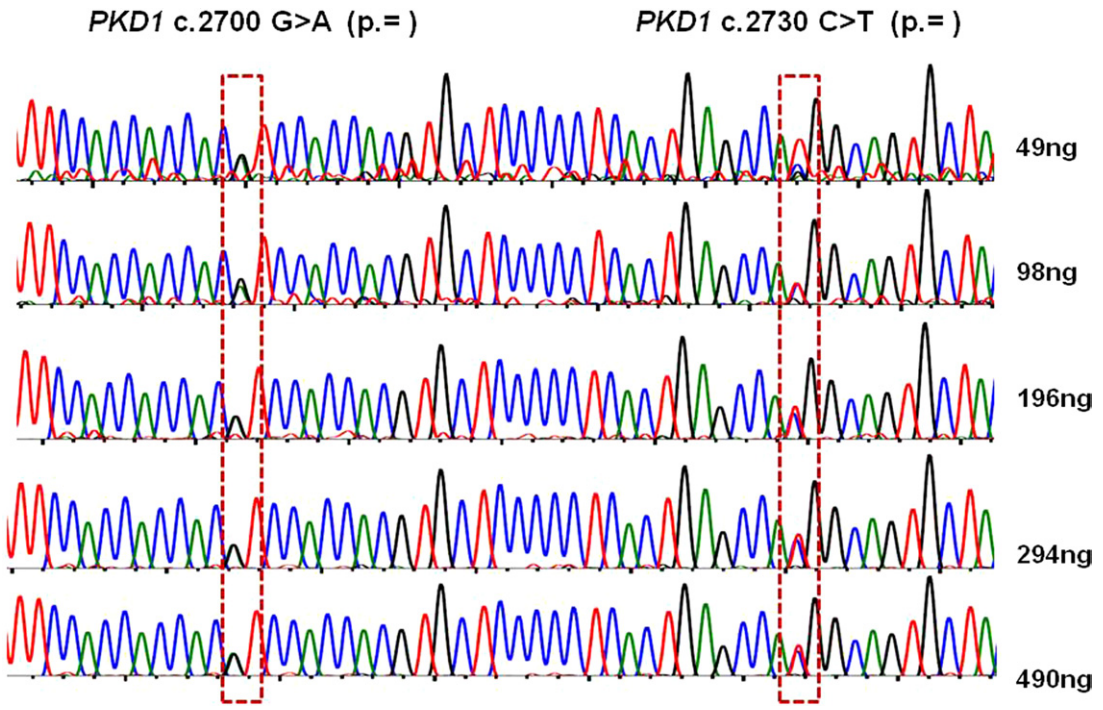

Figure 2. Titration of the amount of input LR PCR DNA in sequencing reactions. The input DNA varied from 49 to 490 ng. An example of an electropherogram of exon11 (fragment Gene8–12) is shown. There are two variants in the regions *PKD1* c.2700 G>A (p. =) and *PKD1* c.2730 C>T (p. =). **Dashed boxes** indicate heterozygous mutation in PKD1.

CIs of sensitivity and specificity were calculated with the VassarStats statistical application (*http://faculty.vassar.edu/lowry/clin1.html*, last accessed May 5th, 2012).

Reference Sequences and Variant Nomenclature

National Center for Biotechnology Information RefSeq sequences were used for reference sequence *PKD1*:

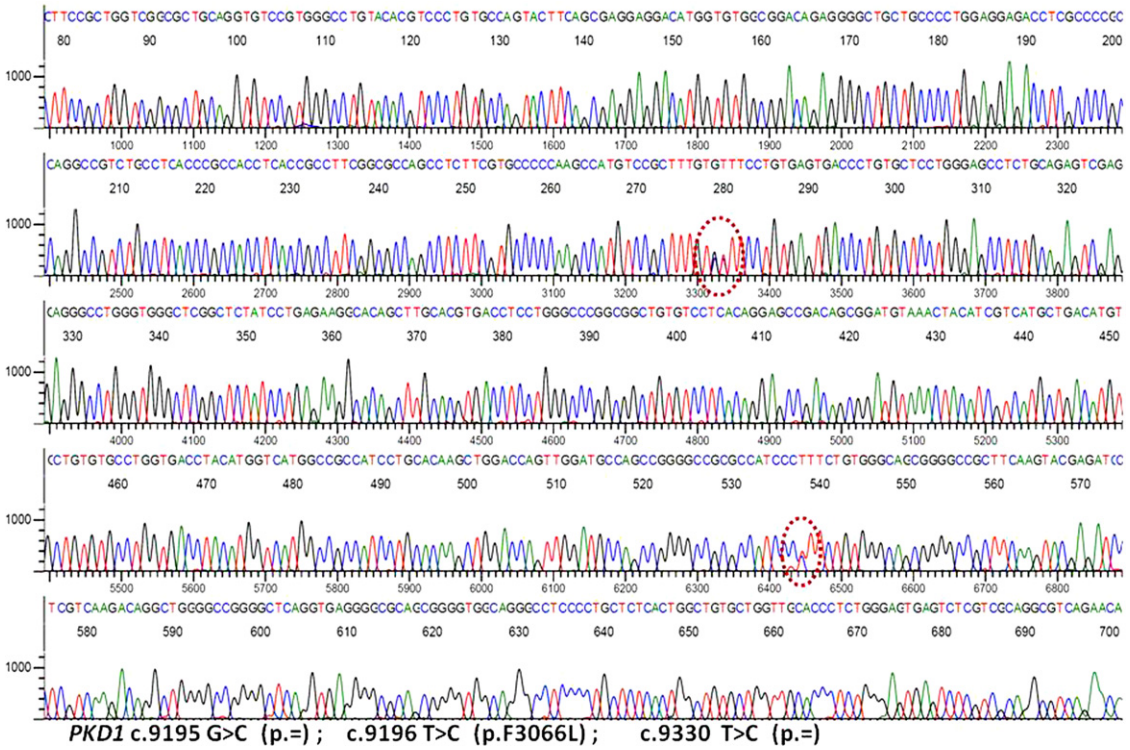

Figure 3. A typical electropherogram of Sanger sequencing using 400 ng of LR PCR product. There is a clean background with low signal to noise, and the three variants [*PKD1* c.9195 G>C (p. =), c.9196 T>C (p.F3066L), and c.9330 T>C (p. =)] in the region could be easily called. **Dashed circles** indicate heterozygous mutation in PKD1.

Table 4. Summary of Heterozygous *PKD1* Gene Variations by Method

Method	SURVEYOR-WAVE	Direct sequencing*	LR PCR sequencing
Missense	14	13	14
Silent	20	17	20
Intronic	8	2	19
Total	42	32	53

*Reference laboratory.

NM_001009944.2. The standard nomenclature recommended by the Human Genome Variation Society (*http://www.hgvs.org/mutnomen*, last accessed January 15, 2012)[13] was used to number nucleotides and name mutations or variants. Sequences were analyzed compared with the *PKD1* gene sequence (accession number: NG_008617.1) and HGs (accession numbers: NG_002797.3, NG_002795.4, NG_002796.4, NG_002800.3, NG_002798.3, and NG_002799.3). All sequence variant descriptions were checked for accuracy using the Mutalyzer 2.0 program (Mutalyzer website; *http://www.mutalyzer.nl/2.0*, last accessed January 15, 2012).

Results

To improve *PKD1* testing and to eliminate the need for nested PCR of the duplicated 5′ region, we used an LR PCR strategy to specifically amplify and directly se-

Table 5. Detailed Genotyping Analysis Results for *PKD1* by Method

Variants					Patients					
					KS9			LD9		
Exon or intron	Nucleotide change	Amino acid change	Type	dbSNP no.	SR-WAVE Seq	Direct Seq	LR PCR Seq	SR-WAVE Seq	Direct Seq	LR PCR Seq
Exon 9	c.1781T>A	p.F594Y	M							
Intron 9	c.1850-4A>G		I	*rs35929659*	X	X	X			
Exon 11A	c.2216A>G	p.Q739R	M	*rs40433*						
Exon 13	c.3063T>C	p. =	S	*rs2369068*	X	X	X			
Exon 14	c.3275T>C	p.M1092T	M	*rs2549677*	X	X	X			
Exon 15A	c.3372C>T	p. =	S		X	X	X			
Exon 15A	c.3375C>T	p. =	S		X	X	X			
Exon 15C	c.4195T>C	p.W1399R	M		X	X	X			
Exon 15D	c.4546G>A	p.A1516T	M		X	X	X			
Exon 15D	c.4665A>C	p. =	S	*rs71385734*	X	X	X			
Exon 15D	c.4674G>A	p. =	S					X	X	X
Exon 15E	c.5172C>T	p. =	S	*rs9935526*	X	X	X			
Exon 15F	c.5611G>A	p.A1871T	M							
Exon 15G	c.5855G>A	p.G1952D	M							
Exon 15I	c.6598C>T	p.R2200C	M							
Exon 17	c.7165T>C	p. =	S	*rs2457533*	X	X	X			
Exon 18	c.7265C>A	p.T2422K	M		X	X	X			
Exon 18	c.7441C>T	p. =	S	*rs2003782*	X	X	X			
Exon 20	c.7708T>C	p. =	S	*rs28575767*	X		X			
Exon 21	c.7913A>G	p.H2638R	M	*rs9936785*	X	X	X			
Exon 24	c.8916C>T	p. =	S		X	X	X			
Intron 24	c.8949-17A>G		I	*rs9928278*	X	X	X			
Exon 25	c.8964G>A	p. =	S					X	X	X
Exon 25	c.9195G>C	p. =	S	*rs9935834*	X	X	X			
Exon 25	c.9196T>C	p.F3066L	M	*rs9925969*	X	X	X			
Exon 26	c.9330T>C	p. =	S		X	X	X			
Exon 35	c.10535C>T	p.A3512V	M	*rs34197769*	X	X	X			
Intron 38	c.11017-5G>T		I							
Exon 44	c.12133A>G	p.I4045V	M	*rs10960*	X	X	X			
Exon 45	c.12176C>T	p.A4059V	M	*rs3209986*	X	X	X			
Exon 45	c.12276A>G	p. =	S	*rs3087632*	X	X	X			
Exon 45	c.12409C>T	p. =	S		X	X	X			
Exon 46A	c.12630T>C	p. =	S	*rs7203729*	X	X	X			
Exon 5B	1119C>T	p. =	S				X, H			X, H
Intron 10	c.2097 + 56C>T		I							X
Intron 11	c.2854-286C>G		I							
Intron 11	c.2854-269T>C		I				X			
Intron 15	c.6915 + 77A>T		I							
Intron 20	c.7863 + 47T>G		I	*rs28610092*	X		X			
Intron 21	c.8016 + 26T>C		I	*rs9934488*	X		X			
Intron 22	c.8161 + 21T>C		I	*rs4786209*				X		X
Intron 22	c.8161 + 38G>A		I	*rs57922772*			X	X		X
Intron 22	c.8161 + 269C>T		I				X, H			
Intron 22	c.8161 + 201_8161 + 220del		I							
Intron 22	c.8161 + 288_8161 + 297del		I				X			
Intron 22	c.8162-262_8162-263_insCT		I							
Intron 23	c.8792-146A>G		I				X			
Intron 39	c.11269 + 128C>T		I	*rs12918803*			X, H			

table continues

Only variants tested by all three methods (SR-WAVE Seq, Direct Seq, and LR PCR Seq) were used to calculate the analytical characteristics of the new LR PCR sequencing assay.

dbSNP, single nucleotide polymorphism database; I, intonic change; M, missense change; S, synonymous change; Seq, sequencing; SR-WAVE, SURVEYOR-WAVE; X, heterozygous change; H, homozygous change; blank cell indicates reference allele.

quence the entire coding region of the gene. By using computational analysis, we designed a set of nine oligonucleotide primer pairs that matched the single-copy gene and rare sequence differences between *PKD1* and the *HGs* (as described in *Materials and Methods*). The relative primer locations and sizes of those fragments are shown in Figure 1. Agarose gel electrophoresis with ethidium bromide staining of the LR PCR products demonstrated specific fragments ranging from 2 to 6 kb, each covering four to seven exons corresponding to the single-copy sequences of *PKD1* (Figure 1). The amount of PCR product required for sequencing will vary and is mostly dependent on the size of the fragment. Careful titration of input DNA demonstrated that at least 300 ng of DNA was required for optimal detection of sequence variations. Figure 2 demonstrates an example of the titration results for the Gene8-12 LR PCR fragment (4 kb) and the two DNA changes identified in this fragment. We chose to use 400 ng of PCR product for sequencing of all LR PCR fragments, which generated high-quality sequencing reactions with high-quality reads of up to 1 kb (Figure 3).

To validate the assay, we selected five unrelated patients with ADPKD who were genotyped by both the reference laboratory and the Cornell Laboratory using direct sequencing and the SURVEYOR-WAVE screening method, respectively. By using the LR PCR sequencing

Table 5. *Continued*

WA9			CM9			MM9		
SR-WAVE Seq	Direct Seq	LR PCR Seq	SR-WAVE Seq	Direct Seq	LR PCR Seq	SR-WAVE Seq	Direct Seq	LR PCR Seq
			X	X	X			
X		X						
X	X	X						
X	X	X						
						X	X	X
						X	X	X
						X	X	X
						X		
X		X			X, H	X		X
					X			
					X			
		X, H			X, H	X		X
		X						
					X			
		X						X

Table 6. Sensitivity and Specificity of the LR PCR Sequencing Method

LR PCR versus direct sequencing				LR PCR versus SURVEYOR-WAVE sequencing			
	Direct sequencing				SURVEYOR-WAVE sequencing		
LR PCR sequencing	Variants (positive)	Reference alleles (negative)	Total	LR PCR sequencing	Variants (positive)	Reference alleles (negative)	Total
Variants (positive)	32	2	34	Variants (positive)	34	0	34
Reference alleles (negative)	0	131	131	Reference alleles (negative)	1	130	131
Total	32	133	165	Total	35	130	165

Compared with the direct sequencing assay (reference laboratory): sensitivity (95% CI) = 32/(32 + 0) = 100% (86.7% to100%), specificity (95% CI) = 131/(131 + 2) = 98.5% (94.2% to 99.7%), and accuracy (95% CI) = (32 + 131)/(32 + 0 + 131 + 2) = 98.8% (95.7% to 99.9%). Compared with the SURVEYOR-WAVE assay: sensitivity (95% CI) = 34/(34 + 1) = 97.1% (83.4% to 99.9%), specificity (95% CI) = 130/(130 + 0) = 100% (96.4% to 100%), and accuracy (95% CI) = (34 + 130)/(34 + 1 + 130 + 0) = 99.4% (96.7% to 100.0%).

method, we identified a total of 53 heterozygous changes (34 exonic and 19 intronic), including all 32 genetic variations reported by the reference laboratory and 41 of 42 variations detected by the SURVEYOR-WAVE sequencing method, indicating that the LR PCR sequencing method has diagnostic performance that is comparable to the screening and direct sequencing methods (Table 4). In addition, this new mutation screening strategy identified 12 intronic changes that were not detectable by the other two methods, as a result of higher intronic coverage. A detailed list of all *PKD1* genetic variations identified in this study is shown in Table 5.

An evaluation of the assay analytical characteristics, using only variants analyzed by all three assays, demonstrated good agreement among the different methods (Table 6). When compared with the standard sequencing method used by the reference laboratory, the LR PCR sequencing assay had a sensitivity of 100% (95% CI, 86.7% to 100%), a specificity of 98.5% (95% CI, 94.2% to 99.7%), and an accuracy of 98.8% (95% CI, 95.7% to 99.9%); compared with the SURVEYOR-WAVE sequencing method, the new assay had a sensitivity of 97.1% (95% CI, 83.4% to 99.9%), a specificity of 100% (95% CI, 96.4% to 100%), and an accuracy of 99.4% (95% CI, 96.7% to 100%) (Table 6).

Discussion

Pseudogene amplification is a particular problem in the molecular diagnosis of ADPKD. Herein, we describe a new *PKD1* mutation screening strategy based on LR PCR amplification of the entire gene, using nine pairs of carefully designed PCR primers. By using this strategy, we have correctly identified, in five randomly selected patients with ADPKD, all gene variations previously identified by the reference laboratory and all except one identified by the SURVEYOR-WAVE screening method described earlier by our laboratory.[12] In addition, this method expands the test coverage to include deep intronic regions, increasing the overall detection rate of genetic variations by 39.6% (from a total of 32 to 53 variants) compared with the reference laboratory.

Humans possess at least six *PKD1* homologues ranging from 18.1 to 29.5 kb, encompassing 33 exons on chromosome 16, significantly complicating genetic analysis and interpretation of genetic variations.[10,14] Recently, we and other laboratories have developed robust LR PCR strategies, using genomic DNA, to enable mutation screening of the reiterated regions of *PKD1*. Similarly, the new strategy described herein broadens the use of this method to the entire gene, including the single-copy regions of *PKD1* (exons 34 to 46), circumventing the need for enzyme digestion of PCR fragments or calibration of PCRs for each of the individual exons. Overall, our method identified all variations reported by the reference laboratory and 97.6% (41/42) of the genetic variations detected by the SURVEYOR-WAVE method, detecting an additional 12 intronic changes that were not detected by the other two methods. We also confirmed that the sequence of these fragments corresponds to a unique sequence, which is *PKD1* and not one of its six homologues.

Direct sequencing is considered the method of choice for mutation detection in many laboratories because screening methods may not detect all sequence or homozygous variations; however, this method requires multiple PCRs and is time-consuming and expensive, limiting the number and size of amplicons primarily to coding regions. In addition, because of variable efficiencies in amplification reactions of target regions, setting up the multiple reactions usually requires intensive optimization, with special attention to GC-rich areas.

A major advantage of the LR PCR method is the elimination of nested PCR for the first 33 exons of the *PKD1*-duplicated region. In our laboratory, this procedure required $1/10^5$ dilutions of the LR PCR fragments, which were then used as templates for the second round of nested PCRs.[11,12] This process was especially susceptible to PCR amplification carryover contamination, potentially leading to false-positive results. Consequently, special precautions were needed, including the use of a second PCR setup chamber and liquid handler. In addition, this method has substantially reduced (by 80%) the number of PCRs. Consequently, the test cost was lower (approximately $1200) and the time to result was significantly less (2 weeks) compared with both the direct sequencing and the SURVEYOR-WAVE screening methods.

PKD1 is highly polymorphic, with an average of 10 variants per individual. Reducing the number of PCRs can reduce the chance for allele dropout during PCR amplifica-

tion, which occurs when a single-nucleotide polymorphism is located in the primer-binding region of an allele.[15]

This new strategy has several other demonstrable benefits over standard sequencing and screening methods. First, it can detect larger genomic rearrangements involving more than one exon that would be missed by amplification of individuals exons. Second, it is suitable for detecting intronic variations important for splicing that may be located as far as 125 nucleotides away from the junctions[16] and are not detected by the standard methods that primarily target the intron-exon junctions of the gene. On average, the LR PCR–based sequencing method covers at least 200 to 300 bp of intronic sequences away from the exon-intron junction. By contrast, the direct sequencing and screening methods used are substantially limited in coverage of intronic sequences to usually not >10 to 50 bp from the exon-intron junction. This is important because intronic changes can induce aberrant splicing by generating alternative splice sites or destroying the branch site; therefore, better coverage of these regions would increase the mutation detection rate.[17–19] Last, it has the ability to detect homozygous changes that are only detected with direct sequencing, but not with other screening methods, such as SURVEYOR-WAVE or denaturing high-performance liquid chromatography, thus significantly improving the overall detection rate of genetic variations.[12,20] Finally, LR PCR methods play an important role in sequencing of repetitive or duplicated regions by next-generation sequencing applications, which cannot be analyzed by the standard genomic amplification or exome enrichment procedures.[21]

The primary limitation of this method is that DNA samples that are partially degraded, such as DNA extracted from paraffin-embedded tissue, can be more difficult to amplify. However, peripheral blood lymphocyte DNA and buccal cell DNA, routinely used for genetic tests, generally work well.[22,23] Another limitation of LR PCR is the requirement for computational evaluation of the duplicated DNA regions and careful primer design to ensure amplification of the single-copy sequences.

In conclusion, the LR PCR sequencing method has several advantages for the detection of *PKD* gene variants, including high sensitivity, improved intronic coverage, faster turnaround time, and lower cost, providing a reliable tool of genetic analysis of complex genes and repetitive sequences. The LR PCR–based strategy can be applied to other genes in which variants have escaped detection, regardless of whether pseudogenes are present.

References

1. Dalgaard OZ: Bilateral polycystic disease of the kidneys: a follow-up of 284 patients and their families. Dan Med Bull 1957, 4:128–133
2. Iglesias CG, Torres VE, Offord KP, Holley KE, Beard CM, Kurland LT: Epidemiology of adult polycystic kidney disease: Olmsted County, Minnesota: 1935–1980. Am J Kidney Dis 1983, 2:630–639
3. Peters DJ, Sandkuijl LA: Genetic heterogeneity of polycystic kidney disease in Europe. Contrib Nephrol 1992, 97:128–139
4. Torres VE, Harris PC: Autosomal dominant polycystic kidney disease: the last 3 years. Kidney Int 2009, 76:149–168
5. Braun WE: Autosomal dominant polycystic kidney disease: emerging concepts of pathogenesis and new treatments. Cleve Clin J Med 2009, 76:97–104
6. Harris PC, Torres VE: Polycystic kidney disease. Annu Rev Med 2009, 60:321–337
7. Dicks E, Ravani P, Langman D, Davidson WS, Pei Y, Parfrey PS: Incident renal events and risk factors in autosomal dominant polycystic kidney disease: a population and family-based cohort followed for 22 years. Clin J Am Soc Nephrol 2006, 1:710–717
8. Blumenfeld JD: Pretransplant genetic testing of live kidney donors at risk for autosomal dominant polycystic kidney disease. Transplantation 2009, 87:6–7
9. International Polycystic Kidney Disease Consortium: Polycystic kidney disease: the complete structure of the PKD1 gene and its protein. Cell 1995, 81:289–298
10. Hughes J, Ward CJ, Peral B, Aspinwall R, Clark K, San Millan JL, Gamble V, Harris PC: The polycystic kidney disease 1 (PKD1) gene encodes a novel protein with multiple cell recognition domains. Nat Genet 1995, 10:151–160
11. Rossetti S, Strmecki L, Gamble V, Burton S, Sneddon V, Peral B, Roy S, Bakkaloglu A, Komel R, Winearls CG, Harris PC: Mutation analysis of the entire PKD1 gene: genetic and diagnostic implications. Am J Hum Genet 2001, 68:46–63
12. Tan YC, Blumenfeld JD, Anghel R, Donahue S, Belenkaya R, Balina M, Parker T, Levine D, Leonard DG, Rennert H: Novel method for genomic analysis of PKD1 and PKD2 mutations in autosomal dominant polycystic kidney disease. Hum Mutat 2009, 30:264–273
13. den Dunnen JT, Antonarakis SE: Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. Hum Mutat 2000, 15:7–12
14. Kirsch S, Pasantes J, Wolf A, Bogdanova N, Munch C, Markoff A, Pennekamp P, Krawczak M, Dworniczak B, Schempp W: Chromosomal evolution of the PKD1 gene family in primates. BMC Evol Biol 2008, 8:263
15. Quinlan AR, Marth GT: Primer-site SNPs mask mutations. Nat Methods 2007, 4:192
16. Majewski J, Ott J: Distribution and characterization of regulatory elements in the human genome. Genome Res 2002, 12:1827–1836
17. Wang K, Zhao X, Chan S, Cil O, He N, Song X, Paterson AD, Pei Y: Evidence for pathogenicity of atypical splice mutations in autosomal dominant polycystic kidney disease. Clin J Am Soc Nephrol 2009, 4:442–449
18. Teraoka SN, Telatar M, Becker-Catania S, Liang T, Onengut S, Tolun A, Chessa L, Sanal O, Bernatowska E, Gatti RA, Concannon P: Splicing defects in the ataxia-telangiectasia gene: ATM: underlying mutations and consequences. Am J Hum Genet 1999, 64:1617–1631
19. Ars E, Serra E, Garcia J, Kruyer H, Gaona A, Lazaro C, Estivill X: Mutations affecting mRNA splicing are the most common molecular defects in patients with neurofibromatosis type 1. Hum Mol Genet 2000, 9:237–247
20. Rossetti S, Chauveau D, Walker D, Saggar-Malik A, Winearls CG, Torres VE, Harris PC: A complete mutation screen of the ADPKD genes by DHPLC. Kidney Int 2002, 61:1588–1599
21. Harismendy O, Ng PC, Strausberg RL, Wang X, Stockwell TB, Beeson KY, Schork NJ, Murray SS, Topol EJ, Levy S, Frazer KA: Evaluation of next generation sequencing platforms for population targeted sequencing studies. Genome Biol 2009, 10:R32
22. Lin J, Kennedy SH, Svarovsky T, Rogers J, Kemnitz JW, Xu A, Zondervan KT: High-quality genomic DNA extraction from formalin-fixed and paraffin-embedded samples deparaffinized using mineral oil. Anal Biochem 2009, 395:265–267
23. Mc Sherry EA, Mc Goldrick A, Kay EW, Hopkins AM, Gallagher WM, Dervan PA: Formalin-fixed paraffin-embedded clinical tissues show spurious copy number changes in array-CGH profiles. Clin Genet 2007, 72:441–447